Middlesex University London

Middlesex University Research Repository

An open access repository of

Middlesex University research

http://eprints.mdx.ac.uk

Fields, Bob ORCID logoORCID: https://orcid.org/0000-0003-1117-1844, Amaldi, Paolo, Wong, B. L. William ORCID logoORCID: https://orcid.org/0000-0002-3363-0741 and Gill, Satinder (2007) Editorial: In use, in situ: extending field research methods. International Journal of Human-Computer Interaction, 22 (1-2) . pp. 1-6. ISSN 1044-7318 [Article] (doi:10.1080/10447310709336952)

Final accepted version (with author's formatting)

This version is available at: https://eprints.mdx.ac.uk/9140/

Copyright:

Middlesex University Research Repository makes the University's research available electronically.

Copyright and moral rights to this work are retained by the author and/or other copyright owners unless otherwise stated. The work is supplied on the understanding that any use for commercial gain is strictly forbidden. A copy may be downloaded for personal, non-commercial, research or study without prior permission and without charge.

Works, including theses and research projects, may not be reproduced in any format or medium, or extensive quotations taken from them, or their content changed in any way, without first obtaining permission in writing from the copyright holder(s). They may not be sold or exploited commercially in any format or medium without the prior written permission of the copyright holder(s).

Full bibliographic details must be given when referring to, or quoting from full items including the author's name, the title of the work, publication details where relevant (place, publisher, date), pagination, and for theses or dissertations the awarding institution, the degree type awarded, and the date of the award.

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Middlesex University via the following email address:

eprints@mdx.ac.uk

The item will be removed from the repository while any claim is being investigated.

See also repository copyright: re-use policy: http://eprints.mdx.ac.uk/policies.html#copy

Editorial:
In-use, In-situ: Extending Field Research Methods

Special issue guest editors:

Bob Fields, Paola Amaldi, William Wong, Satinder Gill

Interaction Design Centre
School of Computing Science (Ravensfield House)
Middlesex University, London, NW4 4BT, UK.

b.fields@mdx.ac.uk, p.amaldi-trillo@mdx.ac.uk, w.wong@mdx.ac.uk, s.gill@mdx.ac.uk

A case for evaluating in use and in-situ

Many authors have argued the need for a broader understanding of context and the situatedness of activity when approaching the evaluation of systems. However, prevailing practice often still tends towards attempting to understand the use of designed artefacts by focusing on a core set of tasks that are thought to define the system. A consequence of such focus is that other tasks are considered peripheral and outside the scope of design and evaluation activities. To illustrate the point, consider the experience, familiar to many of us, of being involved in an evaluation activity where participants provide unstructured qualitative feedback. Irrespective of whether the activity is carried out in a laboratory, in a high fidelity simulation or in a naturalistic setting, participants will frequently volunteer unsolicited feedback about tasks and goals that were not originally within the ambit of the design activity. This unprompted feedback, we suggest, is a cue for the evaluators to pay attention to the relationship between the tool and the practice in which it will be used. In other words a cue to consider the situations in which artefact will be used, the tasks and activities that may be affected by the new system, and so on. These are empirical questions that cannot be answered a priori by the development team, whether the evaluation is taking place in "artificial" or "natural" setting.

Even when context is taken very seriously, there is often an assumption that evaluation is conducted under controlled circumstances, where the evaluators make choices about participants, context, tasks to be performed, and so on. Evaluation in such settings involves several artificial elements: the physical location is not the one that the finished product will ultimately be used in, the test 'users' may or may not be representative of the 'real' users, the tasks carried out in an evaluation may not be those that the product will ultimately be used for, the motivations and other concurrent activities of participants are different in a test situation from those in 'real life', and so on.

Recognising the importance of 'situation', some practitioners have sought to replicate some of the physical surroundings in which things will ultimately be used – for instance by making a usability lab look like a living room. In a similar way, it is common practice in some industries to reflect the complexity of operators' work by constructing highly

sophisticated simulations of the work domain to ensure that evaluations are conducted in a 'realistic' environment.

Scenarios such as these, where fake living rooms or simulations of complex technological phenomena are used to contextualise evaluation activity raise some interesting questions about the degree of 'realism' required for an evaluation to be meaningful, and about the aspects of use that are important enough to preserve in an evaluation set-up.

For instance, it is often held that very 'rough and ready' prototypes can be effective in evaluating new design concepts, and it has been argued that it is precisely the low-fidelity nature of prototypes that allows the evaluation to focus on the most important issues (Rettig, 1994). Another form of compromise is the much quoted claim that in usability testing, most of the important issues will be raised by testing a system with only 5 users (Nielsen, 2000). This reliance on a small number of test users has been disputed (e.g. by Woolrych & Cockton, 2001), but the questions raised are interesting.

Advice is often given to conduct tests in a 'familiar environment', but this is usually a reflection of a concern to place test users in a setting that is free from intimidating influence, rather than of an understanding of the scope of the evaluation effort. Again the issue being not how "realistic" the evaluation setting is, but what are the tasks and relationships that constitute the "focus" on the practice to be considered. Some authors do raise the importance of 'ecological validity'. For instance Benyon et al. (2005) encourage evaluators to consider what aspects of social and physical context may affect use, and to configure evaluation situations to be as "close to the context of use as possible". Once again, the assumption is that evaluation is being carried out in a setting constructed for the purpose of evaluation, and the concern then becomes how to arrange for this setting to be 'ecologically valid'. An alternative agenda, however, arises if we attempt to evaluate not under conditions created by the evaluator, but in the context of use.

It is against this background of concerns that the current set of papers was produced to explore how evaluation might be conducted in the actual settings in which systems will be deployed, "in situ" and actually "in use", rather than in contrived situations divorced from real practice. What motivated the preparation of this special issue (and the earlier workshop from which it grew) was a central concern with the relationship between the way evaluation tests and test situations are configured and the way that work is actually carried out.

The case for more situated forms and methods of assessing designs is made eloquently in the paper by William Gaver and his co-authors. In the context of the design of technologies for use in domestic arenas, they observe that the difficulty of conducting meaningful evaluations is not eliminated when evaluations are conducted in laboratories that are made to look like homes. "The domestic environment doesn't depend purely on the *appearance* of home, after all, but crucially on the situated *activities* that make the home". A similar sentiment is echoed by Seth Chaiklin in relation to the fidelity of systems used in evaluation, by asking "whether fidelity is in relation to the surface appearance .... or the demands of the situation".

The same argument for the importance of in-situ evaluation could apply to many other domains: rather than replicating the physical characteristics of a work system, what matters is assessing a design in the context of the situated activity that takes place.

Lessons Learnt: Themes and dilemmas in the study of technology use in situ

Taken together, the papers highlight a number of themes or issues that begin to emerge when one takes seriously the 'in use, in situ' evaluation agenda, and which are brought into sharper focus when one considers what methods and techniques are appropriate for such evaluations.

1. The trade-off between control and realism

The theme of conducting evaluation in the context of 'real' use is addressed in different ways by all the papers in this issue. The paper by Kjeldskov and Skov opens the discussion by making an interesting distinction between different forms of evaluation. A distinction is made between studies carried out in a laboratory, 'in vitro', and in the field, 'in situ'. Basing their insights on a number of case study evaluations, the authors suggest an intermediate class of evaluation studies, 'in sitro'. Such studies employ high fidelity simulations to achieve a balance between 'realism' on the one hand and 'control' on the other that can be lacking in both lab and field.

2. The need for context in evaluation to expose the unanticipated

The value of exploring technology use in naturally occurring settings through ethnographic or similar field research methods is a theme discussed by several of the papers. Observation of activities 'in the wild', especially over an extended period has the potential to reveal a range of sometimes surprising behaviours that are unlikely to be displayed 'in captivity' in a laboratory setting. For instance Esbjörnsson, Juhlin and Weilenmann describe some of the 'interactional adaptations' that occur as car drivers accommodate mobile phone use into their driving practice. Such adaptations, essential to an evaluation of how a technology functions in practice, while accessible through the kind of ethnographic field research Esbjörnsson et al describe, are unlikely to be displayed in a more conventional laboratory setting. Similarly, Wilson, Galliers and Fone identify a number of unanticipated dimensions of the use of cognitive artefacts in medical work through ethnographic field studies. The authors raise the interesting question of whether carrying out evaluations 'in situ' is sufficient. They argue that observing artefacts 'in use' as part of real work activities reveals aspects of usage that are unlikely to be evident in more conventional user testing where users perform pre-defined tasks, whether 'in situ' or 'in vitro'.

3. Techniques for capturing the situatedness and context of activities

Evaluation is sometimes regarded as a process of understanding how a product will be used, what are the appropriate interactions among which tasks, what the problems in its use are likely to be. Jones et al. consider evaluation in the broader context of design, and

look for ways that employ an understanding of use in the re-design of existing products. The authors make the case for employing a combination of techniques – the interaction trail viewer to give context to data logs of websites visited – to both discover potential new product features, and to evaluate the usefulness and effectiveness of these new features once implemented. Once again, it is argued that a combination of methods used to study how designs function in use, in actual situations of use, is revealing, in this case of potential new features and a sense of their importance to the work being carried out, in ways that more traditional approaches are not.

If the need to evaluate 'in context' is taken as a given for most of the authors, it is observed that some situations of use pose particular problems for evaluation, either because the use setting is hard to access or easy to disturb. Consolvo and her co-workers face the problem of developing ubiquitous technologies for use in the home or on the move. On the one hand it is clear, argue the authors, that such technologies must be evaluated in ways and in settings that reflect the unpredictable ways they are used. On the other hand, conducting such evaluations in the apparent chaos of the real world presents many difficulties. Consolvo et al. describe, through case studies, a number of creative responses to this problem, involving the adaptation of existing methods such as the Wizard of Oz and Experience Sampling for studying use of mobile devices to assist in the collection of data.

Similar problems are encountered by Gaver and his co-authors when they describe the evaluation of 'ludic' designs: artefacts designed to be used in domestic settings in a playful way. Success for such products is focussed far less around whether they support the user in achieving some goal, and is more concerned about the 'openness' of the designs, the ability of the designs to be accommodated into existing practices in new and interesting ways, and whether users are sufficiently engaged with a design to continue to play with it over a period of time. The authors take as their starting point the assumptions that evaluation based around the formulation and testing of hypotheses will miss important features of the way such a design is used. The approach to assessment is therefore based around observing and interpreting the way people encounter the designs over an extended period. In addition to carrying out ethnographic studies, a novel approach to evaluation is proposed, employing 'cultural commentators', who have expertise in explaining and interpreting cultural experiences.

4. Re-creating the social-cultural context of use

Another theme that some of the work reported here touches on, is concerned with the aspects of 'situation' that are relevant when we speak of evaluating 'in situ'. Many of the papers have emphasised the importance of physical, organisational and activity aspects of context. Abdelnour-Nocera, Dunckley and Sharp cast the net somewhat wider in their evaluation of the utility and usability of information systems. Central to their analysis is the way that social context and local culture shapes the perspectives of people who encounter a system. It is shown that perceptions about usability and usefulness are profoundly shaped by, among other things, a person's membership of developer or user cultures.

In a conceptual discussion, Seth Chaiklin picks up a related theme, identifying two approaches to understanding the relation between designed products and the practices they will come to be used in. In the *modular* approach, artefacts can be developed with an understanding of technical possibility and a focus on some pre-defined tasks, but with little regard for social context of use. In contrast, according to the *integrated* approach, a thorough and principled understanding of practice is seen as vital to the successful design and evaluation of systems. The distinction between these two approaches with their emphases on technical possibility and on accommodation within practice, echo the distinction made by Abdelnour-Nocera et al, and reflect a sentiment present in several of the papers in this issue. Chaiklin goes on to offer the theory of activity as a possible underpinning, allowing us to better understand the technology-practice relationship, and setting an agenda for how evaluative studies may be conducted in use and in situ.

In conclusion, as we write this editorial we ask ourselves, "So, what has this volume contributed to our understanding about methods for evaluating technology use in situ?" Incorporating socio-cultural contexts in our evaluations can provide valuable and unexpected insights into the use of technology, and to do this we can develop or adapt techniques to capture the situatedness of these activities, particularly as we trade-off the control we have in full laboratory experiments with the uncertainty and unpredictability associated with everyday situations or field environments.

Bob Fields
Paola Amaldi
William Wong
Satinder Gill

Acknowledgements

We would like to extend our gratitude to the many people who generously gave their time and energy in the production of this issue. In particular, thanks must go to the many reviewers and authors who submitted papers for this issue, and to all those who presented or participated in the lively discussions at the workshop on "In use, in situ: Extending field research methods" held in London in October 2005 (Amaldi *et al.*, 2005). The workshop sparked off ideas that then lead to this special issue, and some of the papers in this issue are developments of work presented at the workshop.

References

Amaldi, P., Gill, S., Fields, B., & Wong, W. (Eds.). (2005). Proceedings of in use, in situ: Extending field research methods.Technical Report IDC-TR-2005-001. Available from www.cs.mdx.ac.uk/research/idc

Benyon, D., Turner, P., & Turner, S. (2005). Designing interactive systems: People, activities, contexts, technologies: Addison Wesley.

Nielsen, J. (2000). Why you only need to test with 5 users. Alertbox - www.alertbox.com, 19 March, 2000. Accessed 21.6.06.

Rettig, M. (1994). Prototyping for tiny fingers. Communications of the ACM, 37(4).
Woolrych, A., & Cockton, G. (2001). Why and when five test users aren't enough. Proceedings of IHM-HCI 2001, Volume 2. Cépadèus Éditions. 105-108.